



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

TESTING THE RESULTS OF THE TEACHING OF SCIENCE.

BY EDWARD L. THORNDIKE,
Teachers College, Columbia University.

The topic which I am asked to discuss is one of enormous complexity. The changes in human beings which result from the teaching of science in schools are real, are measurable, and will some day be defined in units of amount as we now define changes in the rate of a moving body or in the density of a gas. But they include thousands of different elements; they vary with every individual; some of them can be demonstrated only long after school is completed; and at present units and scales in which to state changes in knowledge, power, interest, habits, and ideals are mostly matters of faith. An adequate measurement of the changes wrought in one class by one course in physics would be a task comparable to a geological survey of a state or an analysis of all the materials in this building.

I must also at once confess that I cannot bring you the results of specific investigation of educational achievement in science, but only such suggestions as general experience in measuring human faculty and various educational products can provide.

These suggestions fall naturally into two divisions according as one searches for means of measuring the specific information, skill, interest, and habits added by courses in science or the more general changes in total mental make-up, in, for instance, open-mindedness, accuracy, zest for verification, and the like.

The specific changes are, of course, the easier to measure. Indeed, my first suggestion is that we make scientific use of the measurements that we already make. For example, the regular school examinations are, or should be, careful scientific measures of important changes in our pupils. If we would test our classes with the examinations set by other teachers, have the pupils' work graded by other teachers, and print questions,

work, and grades, we should be making a start toward a real measurement of educational achievement:

If examinations are worth giving at all, they are worth giving, at least occasionally, in such a way as to measure not only how well a pupil has satisfied some particular person, but also what he really is or knows or can do in certain special fields.

We need thousands of significant questions, in each science, thousands of "originals" in physics, chemistry, and biology like the originals of geometry; and above all we need to have thousands of classes tested by outside examiners, as has been done in arithmetic, spelling, handwriting, and geography by Rice, Cornman, Stone, Earhart and Thorndike. If an examination, instead of being a hasty, subjective selection of questions graded still more personally (and alas, how hastily), were made a serious educational measurement, the examination papers of a year would alone give us a large start toward knowledge of what science teaching actually does.

Knowledge may, however, be measured more conveniently than by the examination of note-books, essays, or replies to questions of the ordinary sort. These have the merit of adequacy and richness, but the defects of measuring too many things at once and too indefinitely. Greater uniformity in the use of the test, quickness in scoring it, and freedom from ambiguity in the numerical value assigned can be secured by the exercise of enough ingenuity. I will mention two tests as samples of the many that are possible. The first is an adaptation of a test, devised by Ebbinghaus to measure mental efficiency in general, in filling in words omitted from a passage. From even a hastily devised sample presented here it will be seen that this form of test is scored with reasonable ease. The speed of an individual in selecting words to fill the gaps and the appropriateness of his selections together measure his knowledge. The former is scored with no effort at all and the latter with far less effort than is required to evaluate answers to questions, essays, or experimental works. The paragraphs and omissions therefrom should be arranged with care and improved after trial, but it may be of interest to some of you to compare the ratings obtained in six or eight tests of five minutes each like the following:

A body changing its position in space moves in a certain at a certain in the called acceleration. To change either the or the of a moving requires Suppose a pound of lead to be held at rest 500 feet above the surface of the ocean by a string and the string to be cut. The body will toward the of the beginning to with a of just barely over and reaching at the end of one second a of feet from where it started. In one second the will have from to feet per

The second is a very simple development of so-called association tests which I have used with good success in regular examinations in psychology for a number of years. It needs no other explanation other than a sample.

Write after each of these words some fact which it suggests to you.

acceleration,	gravity,	current,	lever,
density,	expansion,	elastic,	inclined.

As useful means of measuring the interests aroused by the study of science, I suggest records of the books taken from public libraries, of the periodicals chosen in public reading rooms, of the collections gathered and objects constructed by pupils, and a modified form of the test just described, the given words being much less easily provocative of thoughts about facts of science, and being mixed if necessary with words that would call up facts of science only in a person absorbed by scientific interests: The sample I give is left without such padding for disguise.

Write after each of these words some fact which it suggests to you.

work,	time,	wave,	square,	positive,
light,	level,	change,	water,	rate,
pull,	book,	mass,	study,	transform,
gas,	long,	contract,	heat,	law.

This latter test of interest should be varied using pictures of say a man rolling a barrel up a board into a wagon, a lightning flash in the sky, an ordinary (, ??????) scales, and the like, with a similar mixture of "innocent" pictures. Besides words and pictures, actual or described events can be used. If such association tests are to be used to measure interest, they should not be used previously in the form calling definitely for facts about science.

These tests of interest may be used to measure both special interest in particular sciences and general interests, as in fact rather than fiction, knowledge rather than opinion, or verification rather than dispute. Of other means of measuring the general changes wrought by the study of science I will mention only two. The first concerns the power to utilize experience well in thought.

What is needed for this purpose is a series of problems or tasks, relative success with which depends as much as possible upon having power to use experience and as little as possible upon having had certain particular experiences. For example, relative success with the problem, "Which is heavier, a pint of cream or a pint of milk?" is determined largely by ability to select in thought the essential fact that cream rises and to infer its obvious consequence. The data themselves are possessed adequately by all or nearly all pupils alike.

To get such problems I wrote some time ago to one hundred teachers of science, half in universities and half in colleges. I quote some of them:

Rain drops are coming straight down. Will a car standing still or one moving rapidly receive in one minute the greater number of drops on its roof and sides?

Is air drawn up a hot chimney or is it pushed up?

Since it is possible for a person to float in water why is it possible for him to sink?

A cylinder and a cone equal in base and in altitude rest on a plane surface. Which is harder to tip over?

A magnet attracts two iron nails. If the magnet is removed will the nails attract each other?

It is harder to keep your hands clean in the winter than in the summer? Why?

How many surfaces, corners, and edges has a cube?

Which has the greater surface—a cube 10 inches on edge or a sphere 10 inches in diameter.

What is the largest mammal in the world?

Does an iron ball weigh more when it is hot than when it is cold?

If a bottle of gas which is lighter than air be placed with its open mouth upward, will the gas escape from the bottle or will the heavier air press the gas back into the bottle?

Is an incandescent lamp film on fire?

Will a ship that will just barely float in the ocean, float on Lake Erie?

Will a pound of popcorn gain or lose weight or stay the same after it has been popped?

The second means of measuring changes in general power to think is an adaptation of one devised by Professor R. S. Woodworth, in which the pupil picks out from such a series as that below, the statements that are logically absurd, not possibly true. It will be seen that statements could be chosen which would test the power of analysis and of thinking things together in any field of science from the most specialized to the most universal. Following is an example of this form of test.

Put a mark in the margin opposite each of the following sentences which is absurd:

Though armed only with his little dagger, he brought down his assailant with a single shot.

Silently the assembly listened to the orator addressing them.

While walking backwards he struck his forehead against a wall and was knocked insensible.

I saw his boat cleaving the water like a swan.

Having reached the goal, I looked back and saw my opponents still running in the distance.

Offended by his obstinate silence, she refused to listen to him further.

The one-armed cripple was attacked by a dog which seized his wrist, but he pushed it off with the other hand.

With his sword he pierced his adversary, who fell dead.

While threading my way through the crowd, I came suddenly upon an old friend.

The storm which began yesterday morning has continued without intermission for three days.

The dogs pursued the stag through flower gardens in full bloom.

That day we saw several ice-bergs which had been entirely melted by the warmth of the Gulf Stream.

While sharpening his three-bladed knife, my cousin cut his middle finger.

Our horse grew so tired that finally we were compelled to walk up all the hills.

The red-haired girl standing in the corner is taller than any of her older brothers.

A bricklayer fell from a new building quite near our house, and broke both his legs.

The hands of the clock were set back, so that the meeting was surely to close before sunset.

Many a sailor has returned from a long voyage to find his home deserted and his wife a widow.

The two towns were separated only by a narrow stream which was frozen over all winter.

The great advantages of these means of measuring intellectual ability lies in their rapidity and objectivity. If well devised, only two answers are possible, the pupil is measured easily, rapidly, and independently of subjective factors, and his condition is defined in terms of a simple numerical value.

There is no time for me to discuss methods of making, recording, and utilizing these or the hundreds of other equally worthy measurements of educational achievements, that is, of changes produced or prevented in human nature. Nor is this a proper occasion to outline the precautions that are required by the complexity and variability of facts of intellect and character and the absence of well-defined scales with equal units and known zero points, in which to measure facts of intellect and character. For our present purpose it is enough to know that, in spite of difficulties, the measurements can be made, and that a man of science can, if he will, be as scientific in thinking about human beings and their control by education, as in thinking about any fact of nature.